SCIENCE CHINA
Life Sciences

• RESEARCH PAPERS •

July 2011 Vol.54 No.7: 626–633
doi: 10.1007/s11427-011-4189-3

Molecular variation and evolution of the tyrosine kinase domains of insulin receptor *IRa* and *IRb* genes in Cyprinidae

KONG XiangHui[1,2], WANG XuZhen[2] & HE ShunPing[2*]

[1]*College of Life Science, Henan Normal University, Xinxiang 453007, China;*
[2]*Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China*

Received December 3, 2010; accepted May 30, 2011; published online June 23, 2011

The insulin receptor (*IR*) gene plays an important role in regulating cell growth, differentiation and development. In the present study, DNA sequences of insulin receptor genes, *IRa* and *IRb*, were amplified and sequenced from 37 representative species of the Cyprinidae and from five outgroup species from non-cyprinid Cypriniformes. Based on coding sequences (CDS) of tyrosine kinase regions of *IRa* and *IRb*, molecular evolution and phylogenetic relationships were analyzed to better understand the characteristics of *IR* gene divergence in the family Cyprinidae. *IRa* and *IRb* were clustered into one lineage in the gene tree of the IR gene family, reconstructed using the unweighted pair group method with arithmetic mean (UPGMA). *IRa* and *IRb* have evolved into distinct genes after *IR* gene duplication in Cyprinidae. For each gene, molecular evolution analyses showed that there was no significant difference among different groups in the reconstructed maximum parsimony (MP) tree of Cyprinidae; *IRa* and *IRb* have been subjected to similar evolutionary pressure among different lineages. Although the amino acid sequences of IRa and IRb tyrosine kinase regions were highly conserved, our analyses showed that there were clear sequence variations between the tyrosine kinase regions of IRa and IRb proteins. This indicates that IRa and IRb proteins might play different roles in the insulin signaling pathway.

insulin receptor gene, tyrosine kinase domain, Cyprinidae

Citation: Kong X H, Wang X Z, He S P. Molecular variation and evolution of the tyrosine kinase domains of insulin receptor *IRa* and *IRb* genes in Cyprinidae. Sci China Life Sci, 2011, 54: 626–633, doi: 10.1007/s11427-011-4189-3

Insulin and insulin-like growth factor play important roles in growth regulation, cell differentiation and metabolism [1–6]. Their actions are mediated by the corresponding member of the insulin receptor family. The insulin receptor family comprises the insulin receptor (IR), insulin-like growth factors receptors (IGFRs) and the insulin receptor-related receptor (IRR). IR and IGFRs are cross-membrane glycoproteins with similar structures [7,8]. The mature IR comprises two α and two β subunits, which form heterologous $\alpha_2\beta_2$ tetramers connected via disulfide bonds. The IR protein is composed of a binding region in its α-subunit and a transmembrane domain and tyrosine kinase region in its β-subunit [7,8]. The tyrosine kinase activity of the β-subunit can be inhibited by the α-subunit. When insulin binds to the α-subunit, the β-subunit kinase activity inhibitor can be released by phosphorylation and the α-subunit changes its conformation, thereby enhancing tyrosine kinase activity [9]. A combination of insulin and IR results in a cascade of phosphorylation to active downstream regulatory signals. Insulin receptors (IR/IGFR/IRR) belong to the tyrosine kinase receptor family, the kinase activity of which is located in the tyrosine kinase region. Study of the molecular variation and evolution of the tyrosine kinase region of the *IR* gene would help in the understanding of *IR* gene divergence and of different *IR* roles in insulin signaling.

In fish, sequence characteristics and expression patterns

*Corresponding author (email: clad@ihb.ac.cn)

of *IR* and *IGFR* genes have been studied in rainbow trout (*Oncorhynchus mykiss*) [10], coho salmon (*O. kisutch*) [11], turbot (*Scophthalmus maximus*) [10,12], Pacific chum salmon (*O. keta*) [13] and chinook salmon (*O. tshawytscha*) [14]. These *IR* genes contain a conserved tyrosine kinase domain of about 400 bp [11]. *IR*, *IGF-IR* and *IGF-IIR* have also been cloned in *Danio rerio* [1,15]. These studies indicate duplications of *IR* and *IGFR* genes in fish. Toyoshima *et al.* [1] cloned and characterized *IRa* and *IRb* from zebrafish. They also studied their expression patterns and found overlapping functions of the two genes during zebrafish embryogenesis. However, these studies are limited to an individual species or to a few species of fish. It would be of great interest to investigate *IR* gene divergence among closely related fish in certain families by analyzing molecular variation and evolution of tyrosine kinase coding sequences (CDSs) of *IR* genes.

The family Cyprinidae is the largest family of freshwater fish and contains 2010 species in about 210 genera [16]. Cyprinids have a wide distribution and are highly diverse with obvious differences in size, food habits and niches among species. Therefore, they are considered as an ideal group to study functional gene divergence. In this study, molecular variation and evolution of the tyrosine kinase domain of *IRa* and *IRb* were analyzed in the family Cyprinidae. This study aimed to identify variation of the deduced amino acid sequence of the tyrosine kinase domain between IRa and IRb proteins within the family Cyprinidae and to understand the molecular evolution and divergence of *IRa* and *IRb*.

1 Materials and methods

1.1 Sample collection

The nucleotide sequences of *IRa* and *IRb* were determined in 37 representative species from the family Cyprinidae and in 5 species from non-cyprinid Cypriniformes. All specimens were from the collections of the Freshwater Fish Museum of the Institute of Hydrobiology, Chinese Academy of Sciences. The collection location, deposition identification and GenBank accession number of all species are listed in Table 1. Muscle or fin tissue preserved in 95% ethanol was used to extract genome DNA.

1.2 Primer design

Based on the conserved regions of *D. rerio IRa* (AF400271.1; NM_001142672.1) and *IRb* (AF400272.1; NM_001123229), primers were designed and optimized to amplify *Ira* (forward primer 5′-ACGGTYAAYGAGTCK-GCCAGYCT-3′, reverse primer 5′-CYTCRGCCACCAT-GCAGTTC-3′) and *IRb* (forward primer 5′-GMGHGAG-AGRATHGAGTTC-3′, reverse primer 5′-CAGCCACCA-TGCAGTTYCTDG 3′).

Table 1 Species of cyprinid and outgroup taxa investigated in this study

Species	Sampling location	Voucher	GenBank accession No.	
			IRa	*IRb*
Culter alburnus	Wuhan, Hubei	ihbCY0380494	EU009543	EU009588
Elopichthys bambusa	Taoyuan, Hunan	NRMT2286		EU009586
Megalobrama amblycephala	Wuhan, Hubei	ihbCY0305004	EU009544	EU009589
Aristichthys nobilis	Wuhan, Hubei	ihbCYK0411001		EU009592
Squaliobarbus curriculus	Wuhan, Hubei	ihbCY0407001	EU009541	EU009584
Ctenopharyngodon idellus	Wuhan, Hubei	ihbCYK0411003	EU009542	EU009583
Pseudobrama simoni	Taoyuan, Hunan	ihbCY0405361		EU009591
Mylopharyngodon piceus	Wuhan, Hubei	ihbCYK0411004		EU009598
Hemiculter leucisculus	Wuhan, Hubei	ihbCY2603026		EU009587
Ochetobius elongates	Tengxian, Guangxi	ihbCY0108003		EU009585
Xenocypris argentea	Taoyuan, Hunan	ihbCY0405138		EU009590
Aphyocypris chinensis	Wuhan, Hubei	ihbCYK0411005		EU009582
Opsariichthys bidens	Taoyuan, Hunan	NRMT2358	EU009540	EU009578
Saurogobio gracilicaudatus	Guizhou	ihbCY0312012	EU009551	
Saurogobio dabryi	Changyang, Hubei	ihbCY0405136	EU009552	
Gobiocypris rarus	Wuhan, Hubei	ihbCYK0411006	EU009545	EU009581
Pseudorasbora parva	Mengla, Yunnan	ihbCY0312003	EU009549	EU009593
Coreius heterodon	Wuhan, Jinkou	ihbCY0312002	EU009550	EU009594
Hemibarbus maculates	Yunnan			EU009611
Rhodeus ocellatus	Wuhan, Hubei	ihbCYK0411008	EU009547	EU009600
Rhodeus lighti	Wuhan, Hubei	ihbCYK0411009	EU009548	EU009601
Paracheilognathus imberbis	Wuhan, Hubei	ihbCYK0411010	EU009546	EU009599

(*To be continued on the next page*)

(*Continued*)

Species	Sampling location	Voucher	GenBank accession No.	
			IRa	*IRb*
Tanichthys albonubes	Guangdong	ihbCYK0411007		EU009580
Schizothorax oconnori	Tibet	ihbCY0510086	EU009566	
Schizothorax lissolabiatus	Tibet	ihbCY0504193	EU009567	
Schizothorax longibarbus	Tibet	ihbCY0510081	EU009565	
Percocypris pingi pingi	Hejiang, Shichuan	ihbCY0205010	EU009564	EU009596
Gymnocypris waddelli	Qinghai	ihbCY0510092	EU009563	
Acrossocheilus monticola	Hechuan, Chongqing		EU009568	EU009607
Onychostoma macrolepis	Guangdong		EU009570	EU009610
Spinibarbus sinensis	Nanchong, Shichuan	ihbCY0207036	EU009569	EU009608
Ptychidio jordani	Yunnan	ihbCY0308004	EU009561	EU009606
Epalzeorhynchus bicornis	Wuhan, Hubei	ihbCY0505291	EU009562	EU009609
Cyprinus Carpio	Wuhan, Hubei	ihbCYK0411014		EU009597
Puntius semifasciolatus	Mengla, Yunnan	ihbCY0405496		EU009595
Rasbora trilineata	Wuhan, Hubei	ihbCYK0411012		EU009602
Danio myersi	Mengla Yunnan	ihbCY0405411	EU009556	EU009603
Hemimyzon sinensis	Yunnan	ihbCYK0311012	EU009574	
Triplophysa orientalis	Qinghai		EU009573	
Gyrinocheilus aymonieri	Wuhan, Hubei		EU009576	
Myxocyprinus asiaticus	Wuhan, Hubei	ihbCY0305001	EU009577	
Misgurnus anguillicaudatus	Dazhao, Yunnan	ihbCYK0411015	EU009575	EU009612

1.3 DNA extraction, PCR amplification and sequencing

Total genomic DNA was extracted from muscle or fin tissues using standard phenol/chloroform extraction procedures [17]. The target *IRa* and *IRb* sequences were amplified from genomic DNA by PCR. The concentration and purity of DNA was determined using Beckman ultraviolet measurement, and DNA was diluted to 30–50 ng μL^{-1}. 60 μL PCR reactions contained 3 μL DNA template, 6 μL 10×Ex *Taq* PCR buffer, 4.8 μL dNTPs (each at 2.5 mmol L^{-1}, pH 8.0), 1.5 μL each oligonucleotide primer (each at 15 μmol L^{-1}), 0.6 μL Ex *Taq* polymerase (5 U μL^{-1}), and 42.6 μL double-distilled H_2O.

The amplification of *IRa* was carried out with the following conditions: denaturation at 94°C for 4 min, 32 cycles of denaturation at 94°C for 50 s, annealing at 62 or 63°C for 50 s and extension at 72°C for 90 s, and a final extension at 72°C for 6 min. The amplification of *IRb* was performed at an initial denaturation at 94°C for 4 min, 32 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 30 s and extension at 72°C for 60 s, with a final extension at 72°C for 6 min.

The amplified fragments were separated by 1.2% agarose gel electrophoresis, purified using an OMEGA kit (Omega Bio-tek, USA), ligated into a T-tailed pMD18-T vector (Takara, Dalian, China) and transformed into DH5α bacteria. Positive clones were identified using a PCR procedure. Triplicate positive clones carrying the target fragments were sequenced on an ABI3730.

1.4 Sequence analyses

Novel *IRa* and *IRb* sequences have been deposited in GenBank (Table 1). Multiple alignments of sequences were performed using ClustalX (1.83) [18], with a gap-opening penalty of 15.0 and a gap-extension penalty of 3.0. The aligned sequences with a manual correction were used to analyze gene characteristics. Based on CDSs of *D. rerio IRa* and *IRb*, the corresponding CDSs of various fish were determined and extracted from DNA sequence databases to be used in the analysis of molecular evolution and variation. The base composition and substitution of *IRa* and *IRb* CDSs were calculated by MEGA 4.1 [19].

1.5 *IR* gene duplication and divergence

IRa and *IRb* tyrosine kinase domain sequences identified in this study and the corresponding sequences of IR family genes (IR/IGF-IR/IRR) identified in GenBank (Table 2) were used to reconstruct phylogenetic relationships among the members of the IR gene family using the UPGMA method implemented in PAUP4b10 [20]. Bootstrap analysis was carried out with 1000 replicates and a heuristic search. The UPGMA tree was used to analyze molecular evolution of *IRa* and *IRb* in the family Cyprinidae.

1.6 Comparison of d_N/d_S of different clusters

Based on *IRa* and *IRb* CDSs, phylogenetic relationships of the family Cyprinidae were reconstructed using the maximum

Table 2 Sequences in GenBank of IR family genes for *Danio rerio* and other animals

Species	IR family genes	Accession No.
Homo sapiens (human)	*IR*	NM_000208.1
Homo sapiens (human)	*IGF_IR*	NM_000875.2
Homo sapiens (human)	*IRR*	NM_014215.1
Mus musculus (mouse)	*IR*	NM_010568.1
Mus musculus (mouse)	*IGF_IR*	NM_010513.1
Mus musculus (mouse)	*IRR*	NM_011832.1
Rattus norvegicus (rat)	*IR*	NM_017071.1
Rattus norvegicus (rat)	*IGF_IR*	NM_052807.1
Xenopus laevis (African clawed frog)	*IR*	AJ132556.1
Xenopus laevis (African clawed frog)	*IGF_IR*	Z50155.1
Danio rerio (zebrafish)	*IRa*	AF400271.1
Danio rerio (zebrafish)	*IRb*	AF400272.1
Danio rerio (zebrafish)	*IGF_IRa*	AF400269.1
Danio rerio (zebrafish)	*IGF_IRb*	AF400270.1
Branchiostoma lanceolatum (Amphioxus)	*ILPR*	S83394.1
Drosophila melanogaster (fruit fly)	*IR*	U28136
Aedes aegypti (mosquito)	*IR*	U72939.1
Caenorhabditis elegans	*IR*	AC084196.1

parsimony (MP) method, as implemented in PAUP4.0b10. The major-rule consensus tree was achieved through heuristic searching. Bootstrap analysis was conducted using the following parameters: nreps=1000, conlevel=50, search= heuristic.

The aligned *IRa* and *IRb* CDS datasets were used to estimate the ratio of non-synonymous to synonymous substitutions (ω or d_N/d_S) for various branches of the MP tree with the program codeml implemented in PAML3.14 [21]. The following three models were used to evaluate the variable ω rates: (i) One-ratio model (M=0) assigns the same ω rate for all branches; (ii) two-ratio model (M=2), with phylogenetic sense, specifies two or more than two branches to have independent ω rates; (iii) free-ratio model (M=b) assumes that each branch has an independent ω rate.

1.7 Variations of amino acid sequences

Using the protein sequences of *D. rerio IRa* (AAM18903.1) and *IRb* (AAM18904.1) as anchors, *IR* CDSs were translated into amino acid sequences and aligned. The parsimony informative amino acid sites were determined and analyzed using MEGA4.1 software to search for similar amino acid variations among closely related species, and to study the implication of amino acid variation in speciation and molecular evolution.

2 Results

2.1 *IR* gene sequence variations

In the present study, 337 coding nucleotides of *IRa* were aligned (the complete sequence of the tyrosine kinase domain is approximately 400 bp). Of these sites, 252 sites were conserved across all taxa examined, and 85 sites were variable with 60 parsimony informative sites. Base compositions of T, C, A and G were 21.9%, 23.5%, 25.3% and 29.3%, respectively. The transition/transversion ratio was 2.6. For *IRb*, 311 coding nucleotide sites were aligned, with 211 conserved sites, 100 variable sites and 65 parsimony informative sites. Base compositions of T, C, A and G were 22.3%, 23.5%, 26.4% and 27.8 %, respectively. The ratio of transition/transversion was 1.7.

2.2 *IR* gene duplication and divergence in Cyprinidae

In this study, *IRa* and *IRb* genes from fish were clustered into a monophyletic group respectively with 100% nodal supports (Figure 1). This demonstrated that *IRa* and *IRb* genes in fish originated from duplicated *IR* genes, and the daughter genes have evolved into distinct genes under different evolutionary pressure. In addition, this result also showed that *IGF-IR* (*IGF-IRa* and *IGF-IRb*) of *D. rerio* were clustered into one monophyletic clade with 100% nodal support. All members of the vertebrate IR gene family were clustered into one clade; *IRR* and *IGF-IR* genes were grouped into separate monophyletic lineages. Insulin-like protein receptor gene (*ILPR*) from Amphioxus (*Branchiostoma lanceolatum*) was at the basal place within the vertebrate gene tree with 100% nodal support. Invertebrate *IR* genes were grouped outside those of vertebrates at the basal position of the gene tree with high nodal supports.

2.3 Analysis of d_N/d_S in different groups

Based on the *IRa* CDS, a consensus tree of 42 equal parsimony trees was achieved using heuristic searching in parsimony analyses (not shown), with tree length=206, consistency index (CI)=0.5728, retention index (RI)=0.7037, and rescaled consistency index (RC)=0.4031. The family Cyprinidae was grouped into Leuciscini, Barbini and Danionini, and the lineage Barbini was supported with a 62% bootstrap value. Based on the *IRb* CDS, a consensus tree of 657 equal parsimony trees was achieved (not shown), with tree length=300, CI=0.4400, RI=0.5591 and RC=0.2460. Although the Leuciscini and the Barbini were respectively clustered into one lineage, the bootstrap nodal support for each lineage was low.

The tyrosine kinase region sequences of *IRa* and *IRb* were highly conserved, especially among closely related species, and most base substitutions were synonymous. According to the reconstructed MP trees, the ω rates of *IRa* and *IRb* CDSs were estimated for various branches using PAML3.14 (Table 3). In the one-ratio model, the ω rate of *IRa* (0.0292) was slightly higher than that of *IRb* (0.0288). In the two-ratio model, the Leuciscini and the Barbini were assigned independent ω rates. For *IRa*, the ω rates of the

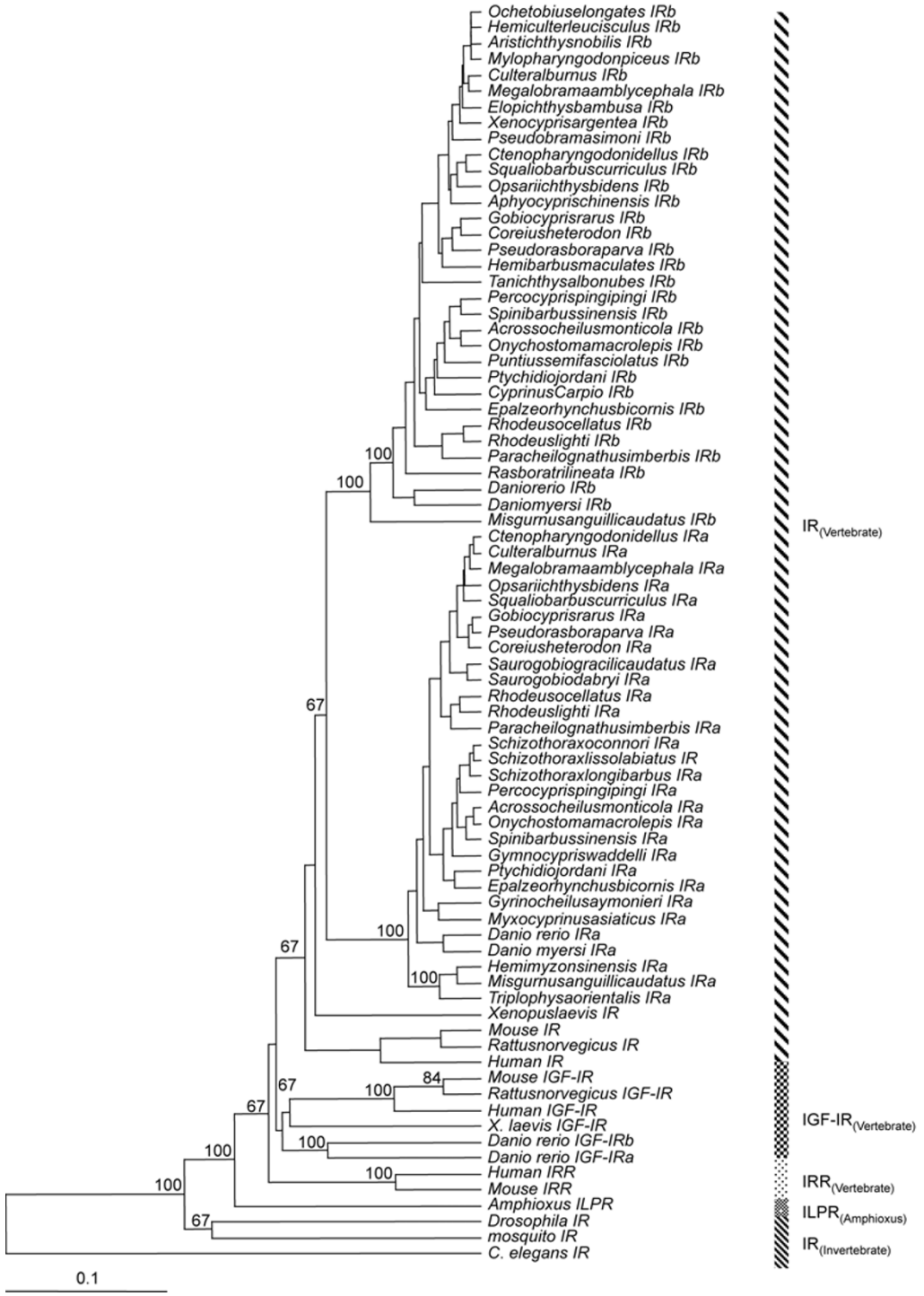

Figure 1 Phylogeny evolution inferred from sequences encoding tyrosine kinase regions of the IR gene family through UPGMA analysis. GenBank accession Nos. of genes are listed in Tables 1 and 2. The number above the branch is the bootstrap score above 50 (only the important nodes are marked).

Leuciscini and the Barbini were 0.0227 and 0.0463, respectively; while for *IRb*, the ω rates of the Leuciscini and the Barbini were 0.0278 and 0.0257, respectively. However, statistical analyses showed that there was no significant difference in the ω rates estimated using the two-ratio model or the free-ratio model compared with the one-ratio model (P>0.05). These data illustrate that there was no significant difference between the Leuciscini and Barbini groups in selection pressure for each gene.

2.4 Amino acid variations of IRa and IRb tyrosine kinase domains

The aligned *IRa* CDS dataset is 337 bp (the first base is the

Table 3 Estimation of ω rates and likelihood values for MP trees[a)]

Model	# ratios	−ln*L*	ω rates (d_N/d_S)
MP tree (*IRa*)			
One ratio	1	1498.74	0.0292
Specific ratio	4	1497.38	0.0227/0.0463/0.0108/0.0311
Free ratio	56	1478.79	Variable by branch
MP tree (*IRb*)			
One ratio	1	2138.54	0.0288
Specific ratio	4	2138.12	0.0278/0.0257/0.0261/0.0415
Free ratio	63	2113.85	Variable by branch

a) Four values in the specified ratio model were the ω rates of Leuciscini, Barbini, Danionini and Outgroup in order.

first position of a codon ACG), which encodes 112 amino acids; the aligned *IRb* CDS dataset is 311 bp (the first base is the third position of a codon CTG), which encodes 103 amino acids. IRa and IRb amino acid sequences were aligned to analyze amino acid variations. The deduced amino acid sequences were highly conserved, and only 10 variant amino acid sites with parsimony information (with phylogenetic sense) were found (without considering gap) (Figure 2), which were situated at sites 9, 25, 37, 42, 57, 64, 65, 79, 83 and 90, respectively. Comparison of amino acid variations between IRa and IRb is showed in Table 4. The similar amino acid variations could be thought of as a representative characteristic to differentiate IRa from IRb. For example, site 9 was arginine (Arg) in most species, while it was serine (Ser) in 9 IRb sequences. For site 42, most IRb sequences (27 species) were serines. For site 57, it was phenylalanine (Phe) in 17 IRa sequences, which were mainly distributed in the Leuciscini. Site 79 was isoleucine (Ile) in 4 IRb sequences, which were in Acheilognathinae. Site 83 was serine in 5 IRb sequences, which were in some species of Daninae. Site 90 was isoleucine in IRa sequences of Schizothoracinae and *Percocypris pingi pingi*.

3 Discussion

In the present study, all members of the vertebrate IR gene

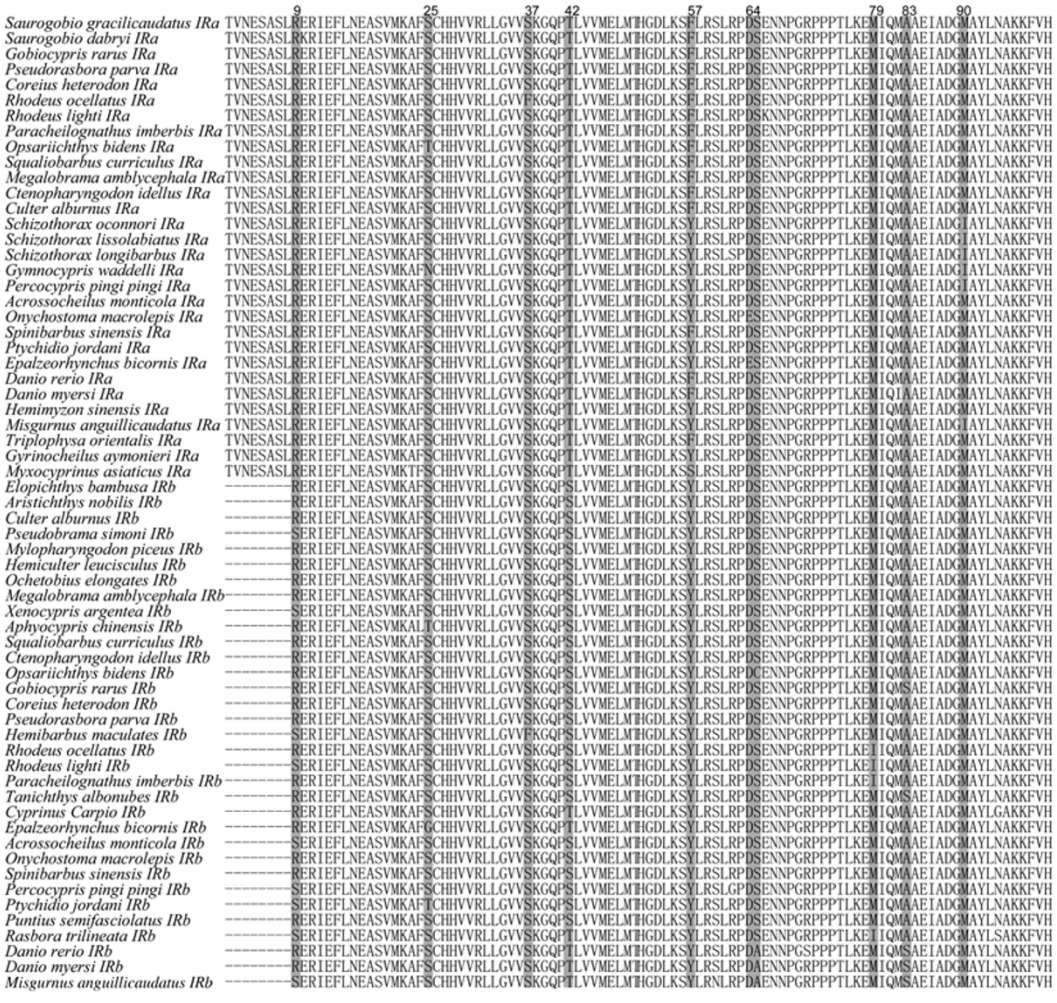

Figure 2 The aligned amino acid sequence dataset based on IRa and IRb. Only variant amino acid sites with parsimony information are marked. The numbers above the sequence dataset represent the position of variant sites.

Table 4 IRa and IRb amino acid variations[a)]

Site	Majority amino acid	IRa variation	IRb variation
9	Arg	0	Ser (9, 27.3%)
25	Ser	Asn (1, 3.3%); Thr (1, 3.3%)	Thr (2, 6.1%), Gly (1, 3%)
37	Ser	Phe (1, 3.3%)	Phe (1, 3%)
42	Thr	0	Ser (27, 81.8%)
57	Tyr	Phe (17, 57%); Ser (1, 3.3%)	0
64	Asp	Glu (2, 6.7%)	0
65	Ser	0	Cys (1, 3%); Ala (3, 9.1%)
79	Met	0	Ile (4, 12.1%)
83	Ala	0	Ser (5, 15.2%)
90	Met	Ile (6, 20%)	0

a) Statistics are based on the amino acid sequence dataset comprising 30 IRa and 33 IRb amino acid sequences. Only variant amino acids are listed in the IRa or IRb columns. The first number in parentheses indicates the sequence number of the variant amino acid; the second number refers to the percent of variant sequences to total sequences.

family were clustered into a lineage. In fish, *IRa* and *IRb* were grouped into one monophyletic clade. This suggests that the daughter genes, *IRa* and *IRb*, have evolved into distinct genes after an ancestral *IR* gene was duplicated. This result is consistent with previous studies on *IR* genes in fish [1,10–15,22]. However, in this study, the divergence of *IR* could provide strong evidence to support the difference of *IRa* and *IRb* in the phylogenetic gene tree. Furthermore, the present study also supported the hypothesis that fish integrated an extra copy of the genome during evolution [22], which has been proposed by studies of *hox* gene duplication in zebrafish and *Fugu* [23]. Gene duplication, followed by functional divergence resulting from mutation and selection, are thought to be critical events that drive biological diversification. This hypothesis of gene function divergence is supported by more and more evidence from studies on functional genes [1,24]. For example, fish-specific duplicated genes *Dmrt2b* and *Dmrt2a* play overlapping roles in the establishment of left-right asymmetry but have divergent functions in somitogenesis [24].

The amphioxus *B. californiensis* has important evolutionary significance in animal phylogeny and evolution and, according to morphological and developmental evidence, is the most primitive species among the chordate animals [15]. An insulin-like protein receptor (ILPR), having a similar function to IR, was separated from amphioxus [25]. *IR* and *IGF-IR* genes of fish might derive from an ancestral gene that was similar to *ILPR* of amphioxus, which duplicated and evolved into *IRa* and *IRb*. This suggestion was also further indicated because all invertebrate animals have only one insulin receptor homologue [26–29]. Amphioxus was showed to be the interim species in the process of evolution from invertebrates to vertebrates.

MP analyses indicated that the Leuciscini and the Barbini were respectively clustered into one lineage within the family Cyprinidae. Some species of the Danioninae (e.g., *D. rerio*, *D. myersi* and *R. trilineata*) were clustered into one group (so called primitive Danionini); however, the inner nodal supports were low. This was because the tyrosine kinase domain of the insulin receptor gene was highly conserved and was, therefore, not suitable for phylogenetic analysis at the subfamily level of fish.

When the Leuciscini and the Barbini were specified to have independent ω rates, the ω rate of the Leuciscini (0.0227) was lower than that of the Barbini (0.0463) for *IRa*; while the Leuciscini (0.0278) was higher than the Barbini (0.0257) for *IRb*. For single *IRa* or *IRb* genes, there is no significant difference between the one-ratio model and the independent ratio model. Therefore, the d_N/d_S heterogeneity of the tyrosine kinase region of *IRa* or *IRb* has not been formed among different groups. Therefore, in fish, after *IR* gene duplication, the daughter genes were functionalized to perform different roles and evolved into two distinct genes (*IRa* and *IRb*). However, because of the important function of these two genes in cell signaling pathways, sequence variation and divergence are subjected to strong functional constraint. Although the strong evidence in this study indicates that *IRa* and *IRb* have evolved in fish into two distinct genes, for the single gene, there was no significant difference of molecular evolution among the different clusters in the family Cyprinidae.

The difference of base composition and variation of CDS sequence between *IRa* and *IRb* indicate that they evolved in different patterns after gene duplication. However, it is well known that the insulin receptor gene, especially the tyrosine kinase region, is subjected to strong functional constraint and selection pressure. Therefore, the amino acid sequence of the tyrosine kinase region was more highly conserved. In the deduced amino acid sequence of IRa and IRb, only 10 amino acid variation sites with parsimony information were found. Of these variation sites, only two sites could indicate the characteristic difference between IRa and IRb proteins (site 42 was serine in 27 IRb sequences; site 57 was phenylalanine in 17 IRa sequences, which was distributed in the Leuciscini, so it could be thought as a characteristic of the Leuciscini). Although variations of the two sites were thought of as characteristic variations, the functional difference of genes resulting from these site variations remains unclear. On the other hand, in the absence of any related research, it is difficult to explain the roles of these variations. However, based on the highly conserved tyrosine kinase regions of IRa and IRb, we have strong reason to believe that amino acid variations could represent important characteristics of gene differentiation and species divergence.

In summary, in the family Cyprinidae, *IRa* and *IRb* have developed into distinct genes under different evolution patterns after *IR* gene duplication, and this suggestion is strongly supported by the phylogenetic relationships of *IRa* and *IRb* in this study. However, tyrosine kinase region CDSs of *IRa* and *IRb* are highly conversed. For a single

gene, there is no significant difference in molecular evolution among the different clusters within the family Cyprinidae. This suggests that *IRa* and *IRb* are subjected to strong functional constraint and evolutionary selection because the insulin receptor plays an important role in cell signaling. Therefore, the amino acid variations between IRa and IRb proteins always act as an important driving force of evolution to propel function divergence. In this study, based on molecular variation of deduced tyrosine kinase domain amino acid sequences, it is postulated that *IRa* and *IRb* might function in different spatio-temporal expression patterns in regulating development and growth in fish. However, this proposal requires confirmation by further studies on expression and physiological functions of IRa and IRb.

We thank Dr. Han YaWei for helpful advice and Hou YuTing for experimental work. This work was supported by the National Natural Science Foundation of China (Grant Nos. 30770299 and 30530120) and the Development Plan of the State Key Fundamental Research of China (Grant No. 2004CB117402).

1 Toyoshima Y, Monson C, Duan C, *et al*. The role of insulin receptor signaling in zebrafish embryogenesis. Endocrinology, 2008, 149: 5996–6005
2 Laustsen P G, Russell S J, Cui L, *et al*. Essential role of insulin and insulin-like growth factor 1 receptor signaling in cardiac development and function. Mol Cell Biol, 2007, 27: 1649–1664
3 Fuss B, Becker T, Zinke I, *et al*. The cytohesin Steppke is essential for insulin signalling in *Drosophila*. Nature, 2006, 444: 945–948
4 Taniguchi C M, Emanuelli B, Kahn C R. Critical nodes in signaling pathways: insights into insulin action. Nat Rev Mol Cell Biol, 2006, 7: 85–96
5 Jones J I, Clemmons D. Insulin-like growth factors and their binding proteins: biological actions. Endocr Rev, 1995, 16: 3–34
6 McMurtry J. Insulin and growth. In: Schreibman M, Scanes C, Pang P, eds. The Endocrinology of Growth, Development and Metabolism in Vertebrates. San Diego: Academic Press, 1993. 249–263
7 Cheatham B, Kahn C R. Insulin action and the insulin signaling network. Endocr Rev, 1995, 16: 117–114
8 LeRoith D, Werner H, Beitner-Johnson D, *et al*. Molecular and cellular aspects of the insulin-like growth factor I receptor. Endocr Rev, 1995, 16: 143–163
9 Saltiel A R, Kahn C R. Insulin signalling and the regulation of glucose and lipid metabolism. Nature, 2001, 414: 799–806
10 Eliès G, Groigno L, Wolff J, *et al*. Characterization of the insulin-like growth factor type 1 receptor messenger in two teleost species. Mol Cell Endocrinol, 1996, 124: 131–140
11 Chan S J, Plisetskaya E M, Urbinati E, *et al*. Expression of multiple insulin and insulin-like growth factor receptor genes in salmon gill cartilage. Proc Natl Acad Sci USA, 1997, 94: 12446–12451
12 Eliès G, Duval H, Bonnec G, *et al*. Insulin and insulin-like growth factor-1 receptors in an evoluted fish, the turbot: cDNA cloning and mRNA expression. Mol Cell Endocrinol, 1999, 158: 173–185
13 Kavsan V, Koval A, Petrenko O, *et al*. Two insulin genes are present in the salmon genome. Biochem Biophys Res Commun, 1993, 191: 1373–1378
14 Wallis A E, Devlin R H. Duplicate insulin-like growth factor-I genes in salmon display alternative splicing pathways. Mol Endocrinol, 1993, 7: 409–422
15 Maures T, Chan S J, Xu B, *et al*. Structural, biochemical, and expression analysis of two distinct insulin-like growth factor I receptors and their ligands in zebrafish. Endocrinology, 2006, 143: 1858–1871
16 Chen Y Y. Fauna Sinica, Osteichthyes, Cypriniformes (II). Beijing: Science Press, 1998. 1
17 Sambrook J, Fritsch E, Maniatis T. Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press, 1989
18 Thompson J D, Gibson T J, Plewniak F, *et al*. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res, 1997, 25: 4876–4882
19 Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings Bioinf, 2004, 5: 150–163
20 Swofford D L. PAUP: phylogenetic analysis using Parsimony, Version 4. Sinauer Associates, Sunderland, Massachusetts. 2002
21 Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci, 1997, 13: 555–556
22 Schlueter P J, Royer T, Farah M H, *et al*. Gene duplication and functional divergence of the zebrafish insulin-like growth factor 1 receptors. FASEB J, 2006, 20: 1230–1232
23 Amorse A, Force A, Yan Y L, *et al*. Zebrafish hox clusters and vertebrate genome evolution. Science, 1998, 282: 1711–1714
24 Liu S, Li Z, Gui J F. Fish-specific duplicated dmrt2b contributes to a divergent function through hedgehog pathway and maintains left-right asymmetry establishment function. PLoS ONE, 2009, 4: 7261–7262
25 Pashmforoush M, Chan S J, Steiner D F. Structure and expression of the insulin-like peptide receptor from amphioxus. Mol Endocrinol, 1996, 10: 857–866
26 Fernandez R, Tabarini D, Azpiazu N, *et al*. The *Drosophila* insulin receptor homolog: a gene essential for embryonic development encodes two receptor isoforms with different signaling potential. EMBO J, 1995, 14: 3373–3384
27 Roovers E, Vincent M E, Van Kesteren E, *et al*. Characterization of a putative molluscan insulin-related peptide receptor. Gene, 1995, 162: 181–188
28 Kimura K D, Tissenbaum H A, Liu Y, *et al*. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. Science, 1997, 277: 942–946
29 Graf R, Neuenschwander S, Brown M R, *et al*. Insulin-mediated secretion of ecdysteroids from mosquito ovaries and molecular cloning of the insulin receptor homologue from ovaries of bloodfed *Aedes aegypti*. Insect Mol Biol, 1997, 6: 151–163